# Repensando el Cociente Intelectual: Un análisis crítico de los fundamentos, limitaciones e implicaciones sociopolíticas de los tests de inteligencia

Ernesto Mascarell Perez
Investigador independiente
mascarell1960@hotmail.com

---

## Resumen

A más de un siglo de la creación del primer test práctico de inteligencia, persiste un notable desacuerdo científico sobre cómo debe describirse la inteligencia humana, si los tests de CI realmente la miden y, en caso contrario, qué es lo que realmente evalúan. El presente artículo ofrece una revisión crítica y sistemática de los fundamentos epistemológicos, metodológicos y éticos de la psicometría de la inteligencia. Se examinan cuatro áreas problemáticas centrales: (1) la naturaleza y estabilidad del constructo "inteligencia" y del factor *g*; (2) el sesgo cultural y ambiental inherente a los instrumentos de medición; (3) las deficiencias metodológicas en los diseños de investigación de heredabilidad; y (4) las consecuencias sociales y éticas de la aplicación de estos tests en contextos educativos, laborales y clínicos. Se argumenta que la pretensión de los tests de CI de medir una capacidad cognitiva general, innata y jerarquizable carece de fundamento empírico sólido y constituye, en gran medida, un artefacto estadístico y cultural. Se concluye con recomendaciones para la investigación futura y la práctica profesional, abogando por modelos alternativos de evaluación del potencial humano.

**Palabras clave:** Cociente Intelectual, tests de inteligencia, factor *g*, sesgo cultural, heredabilidad, psicometría, determinismo biológico, validez.

---

## 1. Introducción

La medición de la inteligencia humana constituye uno de los proyectos más ambiciosos y controvertidos de la psicología científica. Desde que Alfred Binet y

Théodore Simon desarrollaran la primera escala práctica en 1905, los tests de inteligencia se han extendido globalmente, utilizándose en contextos educativos para la identificación de "superdotados" y "discapacitados intelectuales", en el ámbito laboral para la selección de personal, y en el clínico para el diagnóstico de diversas condiciones. Esta ubicuidad contrasta, sin embargo, con la persistencia de debates fundamentales acerca de la naturaleza del constructo que pretenden medir y la validez de sus aplicaciones.

Ken Richardson (2002) planteó con claridad la paradoja central: "A medida que nos acercamos al centenario de la primera prueba práctica de inteligencia, todavía hay poco acuerdo científico sobre cómo debe describirse la inteligencia humana, si las pruebas de CI realmente la miden y, si no es así, qué es lo que realmente miden". Esta incertidumbre no es un mero tecnicismo académico; tiene implicaciones profundas para la movilidad social, la equidad educativa y la justicia distributiva.

El objetivo de este artículo es doble. Por un lado, se propone ofrecer una síntesis crítica de la evidencia empírica que cuestiona los supuestos fundamentales de la psicometría de la inteligencia. Por otro, busca examinar las consecuencias prácticas de la aplicación de estos tests, argumentando que su uso acrítico tiende a reforzar desigualdades estructurales bajo una apariencia de neutralidad científica.

# 2. El problema del constructo: ¿Qué es la inteligencia?

## 2.1 La reificación del factor *g*

El concepto de inteligencia general, o factor *g*, introducido por Charles Spearman a principios del siglo XX, sigue siendo el pilar teórico de la mayoría de los tests de CI contemporáneos. Spearman observó que las puntuaciones en diferentes pruebas cognitivas tendían a correlacionar positivamente —el denominado "manifiesto positivo"— e infirió la existencia de un factor común subyacente a todas ellas.

El paleontólogo Stephen Jay Gould, en su obra *The Mismeasure of Man* (1981), sometió este razonamiento a una crítica fundamental. Gould argumentó que el factor *g* es producto de lo que denominó la "falacia de la reificación": la tendencia a convertir conceptos abstractos en entidades reales. La correlación entre pruebas no demuestra la existencia de una capacidad mental general; puede explicarse igualmente por factores metodológicos, experienciales o sociocognitivos que afectan el rendimiento en cualquier tarea de prueba.

## 2.2 Los problemas de la Teoría CHC

El modelo Cattell-Horn-Carroll (CHC) representa el intento más sofisticado de sistematizar la estructura de las habilidades cognitivas. Este modelo jerárquico de tres estratos —con un factor *g* en el vértice, múltiples habilidades amplias (como la inteligencia cristalizada Gc y la fluida Gf) en un nivel intermedio, y habilidades específicas en la base— ha sido ampliamente adoptado como marco para el desarrollo y la interpretación de tests de inteligencia.

Sin embargo, incluso sus defensores reconocen limitaciones sustanciales. John D. Wasserman (2019) ha señalado que el modelo CHC adolece de "una débil capacidad explicativa, su incapacidad para hacer predicciones comprobables, su enredo con un solo instrumento de medición (el Woodcock-Johnson), la replicación independiente inconsistente de su estructura factorial y, lo más importante, su valor probatorio no demostrado para la identificación e intervención con estudiantes con trastornos específicos del aprendizaje". Wasserman añade que estas limitaciones son inherentes a cualquier teoría psicológica desarrollada o probada principalmente a partir de análisis factoriales.

## 2.3 Alternativas teóricas y su estado empírico

Frente al modelo unitario, han surgido propuestas multidimensionales. La teoría de las inteligencias múltiples de Howard Gardner (1983), que postula la existencia de al menos ocho dominios cognitivos relativamente independientes, ha gozado de gran popularidad en círculos educativos. Sin embargo, su estatus científico es problemático. Una revisión reciente publicada en *Frontiers in Psychology* (2023) concluye que "muchos investigadores han afirmado que la teoría de las inteligencias múltiples de Gardner es un neuromito porque no han visto evidencia que respalde su propuesta de inteligencias independientes basadas en el cerebro". Los estudios factoriales no han demostrado que las inteligencias sean independientes, y la investigación sobre los efectos de la enseñanza basada en MI no ha explorado causas alternativas para los efectos positivos ni se ha realizado mediante métodos científicos estándar.

Por su parte, la teoría de la inteligencia exitosa de Sternberg (1999) propone un modelo triárquico (analítica, creativa y práctica). Sternberg mismo ha criticado tanto a los teóricos del factor *g* —por utilizar rangos restringidos de participantes, materiales o contextos situacionales— como a los defensores de nuevas inteligencias múltiples —por someter sus teorías a pocas o ninguna prueba empírica—, abogando por un término medio que reconozca la naturaleza multifacética de la inteligencia.

Este panorama fragmentado sugiere que, lejos de existir un consenso sobre la naturaleza de la inteligencia, coexisten múltiples marcos conceptuales con diferentes grados de apoyo empírico, ninguno de los cuales puede considerarse definitivo.

# 3. Sesgo cultural y ambiental en la medición

## 3.1 El efecto Flynn y la plasticidad histórica

Uno de los hallazgos más perturbadores para la concepción de la inteligencia como rasgo fijo e innato es el Efecto Flynn: el aumento sostenido de las puntuaciones de CI a lo largo del siglo XX. En 1981, el psicólogo James Flynn observó que las puntuaciones de CI habían aumentado de manera constante durante casi un siglo, con una sorprendente diferencia de 18 puntos a lo largo de dos generaciones. Tras un análisis cuidadoso, Flynn concluyó que la causa era cultural: la sociedad se había vuelto más inteligente en términos de manejar ideas más grandes y abstractas, y había hecho más inteligentes a las personas.

Este hallazgo socava directamente la idea de la inteligencia como una entidad fija e innata. Como señala David Shenk, "esta observación, combinada con la evidencia sólida de que las puntuaciones de CI tampoco son fijas dentro de un individuo, disipa claramente la idea de que la inteligencia es una entidad innata y fija". Los mayores aumentos en las puntuaciones a lo largo del tiempo tienden a asociarse con una reducción del contenido culturalmente orientado en las pruebas de CI, lo que indica que incluso las pruebas diseñadas para ser "libres de cultura" no lo son completamente.

## 3.2 Evidencia de sesgo en pruebas de razonamiento visoespacial

Incluso las pruebas que se presentan como "culturalmente justas", como las Matrices Progresivas de Raven o el test culturalmente justo de Cattell, no escapan al sesgo cultural. Una revisión integradora de 2022 concluyó que "aparecen varios supuestos culturales profundamente arraigados en todas las pruebas de razonamiento visoespacial, hasta el punto de que invalida la visión de tales pruebas como intrínsecamente culturalmente justas y hace imposible extraer conclusiones claras de las diferencias de puntuación media entre grupos étnicos".

Una revisión sistemática de 225 estudios publicada en 2025 identificó diez factores que contribuyen a las diferencias grupales en las pruebas de capacidad

cognitiva, entre ellos: factores biológicos y ambientales, ansiedad ante los tests, actitudes, experiencia y amenaza del estereotipo. Los resultados mostraron que, si bien los factores biológicos y ambientales están relacionados con las diferencias grupales, también lo están varios factores relevantes para el diseño y la administración de las pruebas, "muy notablemente la amenaza del estereotipo, las diferencias en la experiencia y exposición al dominio, y la autoconfianza".

### 3.3 La crítica radical: Los tests de CI como sustitutos de la clase social

Ken Richardson ha llevado la crítica un paso más allá, argumentando que "toda la varianza poblacional en las puntuaciones de CI puede describirse en términos de un nexo de factores sociocognitivo-afectivos que preparan diferencialmente a los individuos para las demandas cognitivas, afectivas y de rendimiento de la prueba —en efecto, que la prueba es una medida del origen de clase social, y no de la capacidad para la cognición compleja como tal". Richardson demuestra cómo estos factores pueden explicar la evidencia correlacional que generalmente se considera que valida los tests de CI, incluidas las asociaciones con el rendimiento educativo, el desempeño ocupacional y las tareas cognitivas elementales.

Esta crítica no implica negar las diferencias individuales en habilidades cognitivas, sino cuestionar que los tests de CI las midan de manera válida y que las diferencias observadas reflejen una capacidad subyacente estable en lugar de desigualdades acumuladas en oportunidades, recursos y experiencias.

## 4. La herencia de la inteligencia: revisión metodológica

### 4.1 El debate sobre la heredabilidad

Uno de los argumentos más recurrentes a favor de la validez de los tests de CI es su supuesta alta heredabilidad, estimada a menudo entre el 50% y el 80%. Un estudio reciente publicado en *Scientific Reports* (2026) reportó una heredabilidad del CI de aproximadamente el 75%, con factores genéticos explicando la mayor parte de la asociación entre CI y estatus socioeconómico (69-98%).

Sin embargo, estas estimaciones deben interpretarse con cautela. La heredabilidad es un parámetro poblacional que varía según el contexto y no implica que las diferencias individuales estén determinadas genéticamente de

manera fija. De hecho, algunos estudios de adopción sugieren que el entorno compartido podría explicar hasta el 50% de la varianza en el CI cuando se corrigen por la restricción de rango.

### 4.2 Interacciones Gen-Ambiente

La investigación sobre interacciones gen-ambiente añade una capa adicional de complejidad. Un metaanálisis de Tucker-Drob y colaboradores (2015) encontró que la interacción entre el estatus socioeconómico y la heredabilidad del CI no era consistente entre países. Mientras que algunos estudios estadounidenses reportaban que la heredabilidad aumentaba con niveles más altos de SES, este patrón estaba "generalmente ausente en estudios fuera de EE. UU. (Inglaterra, Europa, Australia)". Los autores especularon que esta diferencia podría reflejar "peores condiciones para el desarrollo intelectual de individuos de bajo SES en EE. UU. que en otros países".

Un análisis de datos de los Proyectos de Adopción de Texas y Colorado no encontró ninguna interacción heredabilidad x SES sustancial, con estimaciones del tamaño del efecto negativas o cercanas a cero. Esto sugiere que el SES puede no capturar completamente los aspectos ambientales proximales que moderan el CI infantil.

### 4.3 La Cuestión del rango restringido

Un problema metodológico crucial en los estudios de adopción es la restricción de rango (range restriction) del entorno familiar. Las familias que adoptan hijos tienden a ser desproporcionadamente de SES medio-alto, lo que significa que el rango de entornos estudiados no representa la variabilidad total de la población general. Una corrección por esta restricción sugiere que "el entorno compartido podría explicar hasta el 50% de la varianza en el CI". Esto cuestiona seriamente la afirmación de que los factores genéticos explican la mayor parte de la varianza en el CI.

## 5. Implicaciones prácticas y éticas

### 5.1 Validez predictiva en el ámbito laboral

La supuesta validez predictiva de los tests de CI para el desempeño laboral ha sido un argumento clave para su uso en selección de personal. Sin embargo, evidencia reciente cuestiona esta afirmación. Un metaanálisis actualizado de Berry, Lievens, Zhang y Sackett (2024) reveló que "la validez de criterio de las

pruebas GMA [habilidad mental general] ha sido considerablemente sobrestimada debido a correcciones de restricción de rango inapropiadas". Sus resultados llevan a la conclusión de que "excluir las pruebas GMA generalmente tiene poco o ningún efecto sobre la validez, pero disminuye sustancialmente el impacto adverso".

Una revisión crítica publicada en el *Journal of Occupational and Organizational Psychology* (2023) examinó cinco observaciones críticas sobre el uso de pruebas de capacidad cognitiva en la selección para ocupaciones superiores y profesionales: (1) evidencia de impacto adverso en los resultados de las pruebas; (2) la tendencia a posicionar las pruebas de capacidad cognitiva temprano en los procesos de selección de alto volumen; (3) evidencia reciente que desafía la validez metaanalítica de las pruebas de capacidad cognitiva; (4) debilidades en los estudios primarios de validez históricos; y (5) un examen conceptualmente defectuoso de la evidencia de validez diferencial. Los autores concluyen que "la práctica de aplicar pruebas de capacidad cognitiva en estos contextos tiene un fuerte potencial para mantener y exacerbar la desigualdad social en el acceso a ocupaciones y profesiones superiores".

## 5.2 Uso clínico y educativo

Los tests de CI se utilizan en muchos países para determinar la elegibilidad para educación especial, el acceso a servicios de salud y el diagnóstico de discapacidad intelectual. Estas decisiones tienen consecuencias de por vida para los individuos. Sin embargo, dado el sesgo cultural y ambiental documentado en estos instrumentos, existe un riesgo real de que etiqueten erróneamente a niños de entornos desfavorecidos como "discapacitados", mientras que niños de entornos privilegiados son identificados como "superdotados" —no por diferencias en potencial innato, sino por diferencias en oportunidades.

## 5.3 Alternativas: evaluación dinámica

Frente a las limitaciones de los tests estáticos (como los tests de CI tradicionales), han surgido alternativas que se centran en el potencial de aprendizaje en lugar de en logros pasados. La evaluación dinámica, asociada con el trabajo de Reuven Feuerstein y Robert Sternberg, enfatiza "el potencial de aprendizaje en lugar de los logros de aprendizaje pasados". Este enfoque, que sigue un formato de test-entrenamiento-retest, mide la capacidad del individuo para beneficiarse de la instrucción y mejorar con ayuda —una aproximación más cercana a lo que muchos entienden por "potencial" que una medición estática de nivel actual.

Aunque la evaluación dinámica tiene sus propias limitaciones y no está exenta de controversia, representa una dirección prometedora que podría mitigar algunos de los sesgos más problemáticos de los tests tradicionales.

## 6. Conclusión y recomendaciones

La evidencia revisada en este artículo sugiere que los tests de CI, lejos de ser instrumentos neutrales que miden una capacidad cognitiva universal y fija, son artefactos cultural e históricamente situados cuyas puntuaciones reflejan, en medida no despreciable, desigualdades en oportunidades educativas, experiencias socioculturales y recursos materiales. El factor *g*, más que un descubrimiento sobre la arquitectura de la mente humana, parece ser un artefacto estadístico cuya aparente estabilidad oculta una profunda dependencia contextual.

Esto no implica negar la existencia de diferencias individuales en habilidades cognitivas. Implica, más bien, cuestionar la pretensión de que los tests de CI las miden de manera válida y justa, y rechazar la inferencia de que las diferencias observadas reflejan una capacidad general innata jerarquizable.

Sobre la base de esta revisión, se formulan las siguientes recomendaciones:

1. **Para los investigadores:** Abandonar la búsqueda de un único factor *g* universal en favor de modelos multifactoriales y contextualmente situados. Priorizar el estudio de la plasticidad cognitiva y las condiciones que la facilitan sobre la medición de supuestos rasgos fijos.
2. **Para los profesionales (educadores, psicólogos clínicos, seleccionadores):** Ejercer la máxima cautela en la interpretación de las puntuaciones de CI, considerando explícitamente el contexto sociocultural del evaluado y las limitaciones del instrumento. Considerar seriamente el reemplazo de los tests de CI estáticos por enfoques de evaluación dinámica que midan el potencial de aprendizaje.
3. **Para los responsables políticos:** Abstenerse de utilizar los tests de CI como base para decisiones de alto impacto (acceso a educación especial, selección laboral) sin una justificación rigurosa que demuestre su validez para la población específica y su superioridad sobre alternativas menos sesgadas.
4. **Para la formación profesional:** Incorporar una formación crítica sobre la historia, fundamentos y limitaciones de los tests de inteligencia en los planes de estudio de psicología, educación y gestión de recursos humanos, promoviendo una comprensión matizada que reconozca tanto sus posibles utilidades como sus graves limitaciones.

## 7. Declaración de transparencia

El autor declara no tener conflictos de intereses relevantes para el contenido de este artículo. La investigación que sustenta este trabajo se ha basado exclusivamente en fuentes publicadas y accesibles públicamente, y no ha recibido financiación específica.

---

## Referencias


1. Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2), 171-191.
2. Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
3. Gould, S. J. (1981). *The mismeasure of man*. W. W. Norton & Company.
4. Herrmann, E., et al. (2012). Are there geniuses among the apes? *Philosophical Transactions of the Royal Society B*, 367(1603), 2753-2761.
5. Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, 2(2), 194-213.
6. Jensen, A. R. (1980). *Bias in mental testing*. Free Press.
7. McGill, R. J., & Dombrowski, S. C. (2019). Critically reflecting on the origins, evolution, and impact of the Cattell-Horn-Carroll (CHC) model. *Applied Measurement in Education*, 32(3), 216-231.
8. Richardson, K. (2002). What IQ tests test. *Theory & Psychology*, 12(3), 283-314.
9. Richardson, K., & Norgate, S. (2015). Does IQ really predict job performance? *Psychological Science*, 26(2), 145-155.
10. Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Holt, Rinehart & Winston.
11. Sternberg, R. J. (1999). Successful intelligence: Finding a balance. *Trends in Cognitive Sciences*, 3(11), 436-442.
12. Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science*, 27(2), 138-149.
13. Wasserman, J. D. (2019). Deconstructing CHC. *Applied Measurement in Education*, 32(3), 232-248.
14. Wilson, C. J., et al. (2023). The cross-cultural generalizability of cognitive ability measures: A systematic literature review. *Intelligence*, 98, 101751.

15. Woods, S. A., & Patterson, F. (2023). A critical review of the use of cognitive ability testing for selection into graduate and higher professional occupations. *Journal of Occupational and Organizational Psychology*.
16. Cuppello, S., Zibarras, L. D., & Corr, P. J. (2025). Factors related to mean score group differences on cognitive ability tests: A systematic review. *Trends in Psychology*.